Astronomical Imaging with Maximum Entropy

Retrospective and Outlook

Andy Strong
MPE

Interdisciplinary Cluster Workshop on Statistics
Garching, 17-18 Feb 2014
Image reconstruction from incomplete and noisy data

S. F. Gull & G. J. Daniell*

Mullard Radio Astronomy Observatory, Cavendish Laboratory, Madingley Road, Cambridge, UK

Results are presented of a powerful technique for image reconstruction by a maximum entropy method, which is sufficiently fast to be useful for large and complicated images. Although our examples are taken from the fields of radio and X-ray astronomy, the technique is immediately applicable in spectroscopy, electron microscopy, X-ray crystallography, geophysics and virtually any type of optical image processing. Applied to radioastronomical data, the algorithm reveals details not seen by conventional analysis, but which are known to exist.

To avoid abstraction, we shall refer to one example. Starting with incomplete and noisy data by the Backus–Gilbert method a series of maps of radio brightness across the sky, all of which have the data, but have different resolutions and more data alone, there is no reason to prefer any of the observer may select the most appropriate specific question. Hence, the method cannot ‘best’ map of the sky. There is no single suitable for discussing both accurate flux source positions.

Nevertheless, it is useful to have a single...
Imaging with Maximum Entropy

First introduced in astronomy in 1978 by Gull and Daniell, Nature

Radio astronomy : Interferometric data
+ X-ray application

Monkey argument :
A team of proverbial monkeys throws photons at pixels. Each image → process through instrument response, compare with data.

Sort maps in two piles:

- Fits the data
- Does not fit data

Candidate maps (huge number of them!)
Reject (even more of them)

Choose the flattest image
By Maximum Entropy

Colloquially: „the flattest map consistent the data“, (or „smoothest“, or „most conservative“). Should show only features for which there is evidence in the data.
Standard entropy formula:

\[ S = - \sum p_i \log p_i \]

Where \( p_i \) = proportion in pixel \( i \)

But pixels are intensities \( I_i \) so need to replace by

\[ S = - \sum \left( \frac{I_i}{m_i} \right) \log \left( \frac{I_i}{m_i} \right) \]

Where \( m \) is a measure, e.g. the flat map, or average over map.

Can derive from monkey argument: combinatorial formulae for photons in pixels.

Gull & Daniell 1978: maximise \( S \) subject to \( \chi^2 = \) number of data points. (using Lagrangian Multipliers).

In any case guarantees positive image, which e.g. Fourier inversion does not.

Key applications: indirect imaging: image space different from data space.

Nonlinear method, requires efficient algorithms for high-dimensional image space.
Radio galaxy 3C31 1.4 GHz, Cambridge 1-mile Telescope

Fig. 1a. Conventional map of the radio galaxy 3C31. The contour interval is increased by a factor of 10 after the first 10 contours. Negative regions are shown hatched. b, Maximum-entropy map of 3C31, on the same scale as Fig. 1a. The contour interval is increased by a factor of 10 after every set of 10 contours, so that there are 750 of the lowest contours to the peak. The box shows the region covered by Fig. 1c. c, The central region of 3C31, analysed by the maximum entropy-algorithm. The contour levels are the same as in Fig. 1b.

Gull & Daniell 1978
Application to gamma rays

COS-B satellite (from MPE, ESA mission 1975-1982)


**Maximum-entropy image processing in gamma-ray astronomy**

J. Skilling  *Department of Applied Mathematics and Theoretical Physics, Cambridge*

A. W. Strong  *Department of Physics, University of Durham*

K. Bennett  *Space Science Division, ESTEC, Noordwijk, Netherlands*
Maximum Entropy Image using COS-B data
E > 150 MeV
~2000 photons
Maximum Entropy Image using COS-B data
E > 150 MeV
~2000 photons

Figure 4. Maximum-entropy map of the COS-B data in the anticentre region centred on the Crab nebula ($l = 185^\circ, b = -5^\circ$). Contours as for Fig. 3.

- Geminga pulsar
- Crab nebula + pulsar
- CG189+1
- CG176-7
Critique of original Gull & Daniell (1978) formulation of Maximum Entropy imaging:

* Choice of one fit statistic ($\chi^2$) is ad hoc; another criterion gives a different map.

* $\chi^2 = N$ does not consider other acceptable $\chi^2$

* Not applicable to Poisson statistics.

* Just one map, no error estimates.
Evolution of MaxEnt imaging.

1978 Original version

1989 'Classic' = Bayesian: MEMSYS5

1998 'Massive Inference'. Replace pixels by point masses. Not much used for imaging, more for spectroscopy.
'Classical' Maximum Entropy
Gull, Skilling 1989

Bayesian Approach

Full posterior distribution in image space
→ Error estimates on pixels

'Stopping criterion' using hyperprior.

Quantified Maximum Entropy
MemSys5
Users’ Manual
Axiom:

If the proportion of some entity which has a given property is known to be $p$, then the most probable estimate of the proportion in some subclass which has that property is (in the absence of any information connecting the subclass with the property) the same number $p$.

For example, if 30% of kangaroos are left-handed, then the most probable estimate of the proportion of kangaroos in Queensland which are left-handed is also 30% (unless more is known about the handedness of Queensland kangaroos).

Remarkably, the consequence of this apparently weak requirement (Shore and Johnson 1980, Tikochinsky, Tishby and Levine 1984, Gull and Skilling 1984a,b) is that the “best” set of proportions $p_i$ ($i = 1, 2, \ldots, L$) on $L$ a priori equivalent cells must be obtained by maximising the entropy

$$S(p) = - \sum_{i=1}^{L} p_i \log p_i$$

No other function will always give the required uncorrelated form of “best” proportions. This result is slightly restrictive in that proportions must sum to 1, whereas more general positive additive distributions need not. In fact, the only acceptable generalisation of this (to PADs $h$ which need not add to 1) is to select $h$ by maximising the entropy

$$S(h) = \sum_{i=1}^{L} (h_i - m_i - h_i \log(h_i/m_i))$$

(1.1)

where $m_i$ is the measure assigned to cell $i$ (Skilling 1988). In the continuum limit the entropy

PAD = positive additive distribution (e.g. image pixels)
If a team of monkeys throws a very large number $N$ of quanta randomly at the $L$ a priori equivalent cells of a distribution, then the probability of obtaining a particular set $(n_1, n_2, \ldots, n_L)$ of occupation numbers shall be proportional to the degeneracy $N!/n_1!n_2!\ldots n_L!$

**INDUCTION ARGUMENT:**

Of course, we do not suppose that distributions of interest have to be formed in this way; we merely remark that we would like to obtain the right answer in that special case. The consequence of this argument (Skilling and Gull 1989, Skilling 1989) is that $\Phi$ must be of exponential form

Prior on image pixels: 

$$\Phi(S) \propto \exp(\alpha S)$$

Posterior for image pixels: 

$$\Pr(h, \alpha, D) = \Pr(\alpha) \Pr(h | \alpha) \Pr(D | h)$$

$$= \Pr(\alpha) \frac{\exp(\alpha S(h) - \mathcal{L}(h))}{Z_S(\alpha) Z_L}.$$
Posterior for image pixels:

\[
\Pr(h, \alpha, D) = \Pr(\alpha) \Pr(h \mid \alpha) \Pr(D \mid h) \\
= \Pr(\alpha) \frac{\exp(\alpha S(h) - \mathcal{L}(h))}{Z_S(\alpha) Z_L}.
\]

Choice of \(\alpha\) ('stopping criterion').

\[
\Pr(\alpha \mid \alpha_0) = \frac{2\alpha_0}{\pi(\alpha^2 + \alpha_0^2)} \quad (\alpha > 0).
\]

Hyperprior on \(\alpha\)

Maximize the evidence \(\Pr(D \mid \alpha_0) \rightarrow \alpha_0\)
Gaussian approximation about \( h \). The Hessian matrix of \( \alpha S - \mathcal{L} \) is

\[
- \frac{\partial^2}{\partial h \partial h} \left( \alpha S(h) - \mathcal{L}(h) \right) = -\alpha \nabla \nabla S + \nabla \nabla \mathcal{L} \\
= \left[ \mu^{-\frac{1}{2}} \right] \left( \alpha I + \left[ \mu^\frac{1}{2} \right] \nabla \nabla \mathcal{L} \left[ \mu^\frac{1}{2} \right] \right) \left[ \mu^{-\frac{1}{2}} \right] \\
= \alpha \left[ \mu^{-\frac{1}{2}} \right] B \left[ \mu^{-\frac{1}{2}} \right]
\]

where

\[
\left[ \mu \right] = \left[ g \right]^{-1} = (-\nabla \nabla S)^{-1}
\]

\[
B = I + A/\alpha \\
I = \text{identity matrix, and } A = \left[ \mu^\frac{1}{2} \right] \nabla \nabla \mathcal{L} \left[ \mu^\frac{1}{2} \right].
\]

Hence the Gaussian approximation to (2.1) is

\[
\Pr(h, \alpha, D) = \Pr(\alpha) \frac{\exp(\alpha S(\hat{h}) - \mathcal{L}(\hat{h}))}{Z_S Z_\mathcal{L}} \exp \left( -\frac{\alpha}{2} (h - \hat{h})^T \left[ \mu^{-\frac{1}{2}} \right] B \left[ \mu^{-\frac{1}{2}} \right] (h - \hat{h}) \right).
\]

Maximize \( P(\alpha | D) \) -> \( -2\alpha S(\hat{h}) = G, \quad G = \text{trace}((\alpha B)^{-1} A) \)

Using the eigenvalues \( \lambda \) of \( A \) to write

\[
G = \sum \frac{\lambda}{\lambda + \alpha}
\]

Information in image ~ number of „good“ measurements in data
Figure 2.1: Maximum entropy trajectory.
Figure 2.1: Maximum entropy trajectory.
Gull & Skilling 1989

Data

Maximum Entropy Deconvolution
Classic Maximum Entropy: some troubles!

Stopping criterion led to overfitting the noise.

Reason: no pixel-to-pixel correlations in entropy formula!

Introduced 'Intrinsic Correlation Function' in MEMSYS5 package.

Image = convolution of 'pre-image' with a kernel

Induces a smooth image – looks nicer!

But this is ad-hoc, which is what was supposed to be avoided!

Still, in astronomy pixels are not necessarily correlated.

galaxy - star - galaxy … do not know about each other necessarily. So this trick is not really needed.
COMPTEL
Compton Telescope on NASA Gamma Ray Observatory
1 – 30 MeV

Ideal MaxEnt application since very broad response.

Cannot make images directly by just binning photons
(e.g. for as Fermi-LAT)
Double Compton scattering: scatter-angle is function of measured energy loss $E_1$ and $E_2$. Each photon defines a circle on sky, different radius for every photon!

Incomplete absorption in lower detectors:
$E_2$ not measured exactly $\rightarrow$ circles are broadened to annuli
If just one source, intersection of circles gives position.
For more general case, multiple sources, extended emission
intersection circle method will not work.

Need *indirect imaging* method.
COMPTEL Maximum Entropy imaging

Ideal application since very broad response, 
Cannot make images directly by just binning photons 
(e.g. like Fermi-LAT)

First application of Classic MEM in gamma ray astronomy. 
MEMSYS5.

Cray - early application of supercomputer, 240 CPUs.

Problems: high background, not handled by package, some tweaking required.

Produced standard skymaps of $^{26}$Al line, and continuum.
CGRO/ COMPTEL
MeV continuum

1 – 3 MeV

3 – 10 MeV

10 – 30 MeV
Error bars using Classical MaxEnt

1–3 MeV Cycle 1–6 SKYMO5/Cray T3E

3–10 MeV Cycle 1–6 SKYMO5/Cray T3E

10–30 MeV Cycle 1–6 SKYMO5/Cray T3E
LS 5039 – the counterpart of the unidentified MeV source
GRO J1823-12

W. Collmar¹, S. Zhang²

¹ Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, D-85748 Garching
² Key Laboratory for Particle Astrophysics, Institute of High Energy Physics, Beijing 100049, China

LS5039
X-ray binary
Microquasar candidate
now identified:
periodic signal

**Fig. 8.** The orbital light curve of LS 5039 in the 10-30 MeV COMPTEL band for the sum of all data. The lightcurve is folded with the orbital period of ~3.9 days and given in phase bins of

arXiv:1402:2525

in press
Maximum Entropy Skymaps

EGRET Cycle 1-2  > 100 MeV

COMPTEL Cycle 1-5  10-30 MeV

Shows that COMPTEL map is probably quite reliable.
Fermi-LAT 25-40 MeV

PRELIMINARY

meets

COMPTEL 10-30 MeV
$^{26}\text{Al}$
Inviting to return to this: far more powerful computers and convolution algorithms.

No follow-up mission foreseen at present in MeV range!
Despite immense discovery potential.

COMPTEL analysis still active at MPE! (Werner Collmar)
INTEGRAL / SPI 2002-2016 (at least)

Coded mask with multiple pointings, hence direct deconvolution not possible.

Usually analysed by model-fitting methods: limited by model.

For real skymaps, \textit{indirect imaging} required.

Challenging: large and time-dependent instrumental background, Has to be determined by model fitting first, in current implementations

Huge amount of data now available, promising future for imaging.
Cornelia (‘Trixi’) Wunderer (now at DESY)  
PhD thesis 2002: INTEGRAL / SPI laboratory mask study  
Engineering model of SPI mask  
(heavy object!, always as demo at Tag der offenen Tür)  

Maximum entropy method for radioactive sources.  

Good resolution.
Figure 7.10: spiskymax reconstruction of two $^{22}$Na sources in the 1274 keV line using 5-point-dithered "BIN" data. Resulting images are shown for source separations of 0.5° to 3.0°.
Fig. 1. *spiskymap* image of the Cygnus region in energy range 200–400 keV, using Performance Validation Phase SPI data.


Fig. 2. *spiskymap* images of the inner Galaxy in energy ranges 18–40 keV (upper) and 40–100 keV (lower), using the first cycle of GCDE SPI data. Sources visible include 4U1700-377 \((l = 347.8, b = +2.2\)), H1741-322 \((l = 357.1, b = -1.6\)), 1E1740.7-2942 \((l = 359.1, b = -0.1\)), Sco X-1 \((l = 359.1, b = +23.8\)), GS1826-238 \((l = 8.9, b = -5.3\)), GRS1915+105 \((l = 45.4, b = -0.2\)).

**Maximum Entropy imaging with INTEGRAL/SPI data**

A. W. Strong*
SPI maximum entropy skymaps

18-143 keV

143-268 keV

268-393 keV

393-518 keV

508-514 keV

positronium

positron annihilation line: 508-514 keV; MaxEnt Strong et al. 2004
INTEGRAL / SPI
Maximum Entropy image
2006: ~3 years data, ~20000 pointings

20-25 keV

Credit: Xiaoling Zhang
INTEGRAL / SPI
511 keV line, 10 years data, preliminary image
\(^{26}\)Al, 1.8 MeV, 10 years data
Preliminary image
Other applications of Maximum Entropy in astronomy:

WMAP : included in standard map products

PLANCK : among the methods used

(NB EVLA / ALMA / CASA Package – CLEAN is still the standard )
The future for MaxEnt imaging?

MEMSYS5 limitation: Hessian approximation to likelihood function.

Modern algorithms (MCMC etc) would allow full exploration of posterior image space, without Hessian approximation for likelihood function.

Eventually probably overtaken by more recent advances like IFT / NIFTY.
Outlook

* Maximum Entropy imaging for Fermi – LAT? Not yet! But very enticing because broad PSF and high statistics at low energy → deconvolution. Now down to 30 MeV, so overlap with COMPTEL.

* COMPTEL - well worth revisiting, increased computer power and algorithms, No mission in MeV in sight, so valuable heritage data at these unexplored energies.

* eRosita: PSF strongly varying over field, many scans, hence candidate for Maximum Entropy for extended sources.

* Looking for collaborateurs on any of these topics – also with more advanced methods!